Accuracy of Hospital Report Cards Based on Administrative Data

Laurent G. Glance, Andrew W. Dick, Turner M. Osler, and Dana B. Mukamel

Context. Many of the publicly available health quality report cards are based on administrative data. ICD-9-CM codes in administrative data are not date stamped to distinguish between medical conditions present at the time of hospital admission and complications, which occur after hospital admission. Treating complications as pre-existing conditions gives poor-performing hospitals “credit” for their complications and may cause some hospitals that are delivering low-quality care to be misclassified as average- or high-performing hospitals.

Objective. To determine whether hospital quality assessment based on administrative data is impacted by the inclusion of condition present at admission (CPAA) modifiers in administrative data as a date stamp indicator.

Design, Setting, and Patients. Retrospective cohort study based on 648,866 inpatient admissions between 1998 and 2000 for coronary artery bypass graft (CABG) surgery, coronary angioplasty (PTCA), carotid endarterectomy (CEA), abdominal aortic aneurysm (AAA) repair, total hip replacement (THR), acute MI (AMI), and stroke using the California State Inpatient Database which includes CPAA modifiers. Hierarchical logistic regression was used to create separate condition-specific risk adjustment models. For each study population, one model was constructed using only secondary diagnoses present at admission based on the CPAA modifier: “date stamp” model. The second model was constructed using all secondary diagnoses, ignoring the information present in the CPAA modifier: the “no date stamp model.” Hospital quality was assessed separately using the “date stamp” and the “no date stamp” risk-adjustment models.

Results. Forty percent of the CABG hospitals, 33 percent of the PTCA hospitals, 40 percent of the THR hospitals, and 33 percent of the AMI hospitals identified as low-performance hospitals by the “date stamp” models were not classified as low-performance hospitals by the “no date stamp” models. Fifty percent of the CABG hospitals, 33 percent of the PTCA hospitals, 50 percent of the CEA hospitals, and 36 percent of the AMI hospitals identified as low-performance hospitals by the “no date stamp” models were not identified as low-performance hospitals by the “date stamp” models. The inclusion of the CPAA modifier had a minor impact on hospital quality assessment for AAA repair, stroke, and CEA.

Conclusion. This study supports the hypothesis that the use of routine administrative data without date stamp information to construct hospital quality report cards may result

in the mis-identification of hospital quality outliers. However, the CPAA modifier will need to be further validated before date stamped administrative data can be used as the basis for health quality report cards.

Key Words. Administrative data, quality of care, measurement, and reporting systems

Improving the quality of health care has become the centerpiece of health care reform in this country. According to the landmark report from the Institute of Medicine, “between the health care we have and the care we could have lies not just a gap, but a chasm” (Crossing the Quality Chasm 2001). As many as 98,000 patients die each year as a result of preventable medical errors (To Err Is Human: Building a Safer Health System 2000). A recent study from the RAND corporation found that Americans receive only 55 percent of recommended care for a group of 30 acute and chronic conditions (McGlynn et al. 2003). These structural problems are even more striking given that “from a technical and scientific standpoint, the capabilities of the nation’s health care system are extraordinary” (Blumenthal 1996). Nevertheless, in practice, many Americans receive substandard and low-quality care (Crossing the Quality Chasm 2001; McGlynn and Brook 2001). Widespread problems in quality of care may not be solvable without undertaking major and systematic changes in the health care system (Chassin and Galvin 1998).

Performance measurement is a fundamental component of health care reform. Health care researchers have formulated a research agenda to bridge the “quality chasm” identified by the Institute of Medicine (IOM) (Fernandopulle et al. 2003; McGlynn et al. 2003; Shaller et al. 2003). A core part of this strategy is to develop a national quality measurement and reporting system, as mandated by Congress (McGlynn 2003). Critical to accomplishing this task is the need to develop “better measures of case-mix and severity of disease” (Leatherman, Hibbard, and McGlynn 2003) in order to improve performance measurement (Fernandopulle et al. 2003).

Address correspondence to Laurent G. Glance, M.D., Associate Professor, Department of Anesthesiology, University of Rochester School of Medicine, 601 Elmwood Avenue, Box 601, Rochester, NY 14642. Andrew W. Dick, Ph.D., Senior Economist, is with the Rand Corporation, Pittsburgh, PA. Turner M. Osler, M.D., Professor, is with the Department of Surgery, University of Vermont College of Medicine, Burlington, VT. Dana B. Mukamel, Ph.D., Professor and Senior Fellow, is with the Department of Medicine, Center for Health Policy Research, Irvine, CA.

One of the major obstacles to the creation of a national quality measurement and reporting system is the absence of computerized and readily available clinical data. The slow adoption of the electronic health record (Hersh 2004) has made clinical databases the exception as opposed to the norm. Accurate data are the foundation of severity adjustment and outcomes measurement. In the absence of widely available clinical databases, many of the existing health quality report cards are based on administrative data. The important role that administrative data can play in quality measurement in the current environment has been recognized by the IOM: “Administrative data, such as Medicare claims, represent one of the most practical and cost-effective data sources on selected components of health care quality available today” (Hurtado, Swift, and Corrigan 2001).

However, the majority of administrative datasets, including Medicare, Medicaid, and most of the state databases, have a very serious limitation—I CD-9CM codes do not distinguish between preexisting conditions and conditions that developed following hospital admission (complications). Treating complications as preexisting conditions will give poor-performing hospitals “credit” for their complications and may cause some hospitals that are delivering low-quality care to be classified as average- or high-performing hospitals (Jollis and Romano 1998). Only administrative datasets from California and New York State contain a condition present at admission (CPAA) modifier for each ICD-9-CM code that makes it possible to distinguish between preexisting conditions and complications. Previous work by our group suggests that adding date stamp information to administrative data may lead to the more accurate identification of preexisting conditions (Glance et al. 2005). In principle, one would expect the CPAA modifier to greatly improve the value of administrative datasets for constructing quality report cards. It is unknown, however, whether this additional information would make a difference and whether the investment that would be required to provide this information on a national scale would be justified.

To gain a better understanding of the impact of date stamping on quality measurement, we conducted a study to assess whether hospital mortality ranking is significantly changed by the inclusion of CPAA modifiers (henceforth referred to as “date stamp”) in administrative data. This study assesses the validity of existing data sources for measuring quality and explores ways in which the validity of these data can be improved (Leatherman, Hibbard, and McGlynn 2003). We used data from the California State Inpatient Database (SID) to examine the impact of date stamping on hospital ranking for patients admitted for each of seven major surgical procedures

or medical conditions: coronary artery bypass grafting (CABG), coronary angioplasty (PTCA), carotid endarterectomy (CEA), abdominal aortic aneurysm (AAA) repair, total hip replacement (THR), acute myocardial infarction (AMI), and stroke.

METHODS

Data Source

This analysis is based on the 1998–2000 California State Inpatient Database (SID) obtained from the Healthcare Cost and Utilization Project, a federal–state–industry partnership sponsored by the Agency for Healthcare Research and Quality. It contains 100 percent of the state’s inpatient discharge records. The California SID has ICD-9-CM coding slots for up to 30 diagnoses. With the exception of E-codes, each ICD-9-CM code is modified using a CPAA field to indicate whether a diagnosis was present at admission.

Seven study populations were defined using ICD-9-CM codes: CABG, PTCA, CEA, AAA repair, THR, AMI, and stroke (Table 1). These procedures and diagnoses were selected because they constitute an important group of moderate-to-high risk surgical and medical conditions. The study was limited to adult patients (age ≥ 18). Hospitals that set all CPAA modifiers equal to one were excluded from the analysis (62 patients). The final dataset consisted of 648,866 patients in 394 hospitals.

Table 1: ICD-9-CM Codes for Study Populations

	<i>Number of</i>		<i>ICD-9-CM Codes</i>	<i>Disease Staging</i>
	<i>Patients</i>	<i>Hospitals</i>		
Coronary artery bypass grafting	84,656	123	36.10–36.39	CVS-01
Coronary angioplasty	149,375	162	36.01, 36.02, 36.05, 36.06	
Carotid endarterectomy	35,374	311	38.12	
Abdominal aortic aneurysm repair	8855	301	38.34, 38.44, 38.64	
Total hip replacement	74,663	360	81.51, 81.52	
Acute myocardial infarction			410.0, 410.01, 410.1, 410.11 410.2, 410.21, 410.3, 410.31 410.4, 410.41, 410.5, 410.51 410.6, 410.61, 410.7, 410.71	Stroke
	170,831	378	410.8, 410.81, 410.9, 410.91	
			433.01, 433.11, 433.21, 433.31	
			433.81, 433.91, 434.01, 434.11,	
	125,112	389	434.91, 436	

Validation of the CPAA Field

The CPAA field has not been validated using chart abstraction or clinical data. An initial validation study performed when the CPAA modifier was first added to CA administrative data (1996) showed only modest agreement with reabstracted data in patients with community acquired pneumonia¹ (Report for the California Hospital Outcomes Project Community-Acquired Pneumonia 1996). However, internal validation studies performed by the California Office of Statewide Health Planning and Development (OSHPD) suggest that the coding accuracy of the CPAA modifier has improved greatly since it was first introduced.²

We performed an exploratory analysis to assess the validity of the CPAA field using the study populations. We first identified ICD-9-CM codes that were likely to represent either chronic conditions (e.g., valvular disease, essential hypertension, old myocardial infarction) or complications (e.g., post-operative shock, cardiac complications, respiratory complications). For ICD-9-CM codes that most likely code chronic conditions, the CPAA field indicated that these secondary diagnoses were present at admission 99 percent of the time (Table 2). For ICD-9-CM codes that designate secondary diagnoses likely to represent complications, the CPAA field indicated that these secondary diagnoses were present at admission in 12 percent of the cases (Table 2). The finding that the CPAA field identifies some ICD-9-CM "complication" codes as being present on admission may be explained by patients readmitted with complications from a prior admission or patients who had prior outpatient procedures.

Two further exploratory analyses were performed to evaluate the uniformity of the application of the CPAA modifier at the hospital level. First, we examined the distribution of the CPAA modifier for ICD-9-CM codes that identify chronic conditions by calculating the proportion of chronic codes identified as present at admission for each hospital. These distributions are shown in Figure 1 for each of the seven study populations. The CPAA modifier appears to be applied to chronic conditions in a uniform manner by the study hospitals. Over 95 percent of chronic conditions were coded as present at admission by all of the hospitals.

Second, we examined the distribution of the CPAA modifier for ICD-9-CM complication codes (Figure 2). We only included the surgical and angioplasty patients in this second analysis because we assumed that conditions coded as complications in patients admitted with either stroke or AMI would be more likely to identify conditions from a previous admission than for

Table 2: Proportion of Chronic Conditions and Complications Coded by the CPAA Modifier as Present at Admission or Not Present at Admission

	ICD-9-CM	n	Present at Admission (%)	Not Present at Admission (%)	Missing (%)
<i>Chronic conditions</i>					
Rheumatic fever w/ heart involvement	391x	43	90.70	9.30	0.00
Chronic rheumatic pericarditis	393	4	100.00	0.00	0.00
Diseases of mitral valve	394x	606	98.35	1.32	0.33
Diseases of the aortic valve	395x	89	100.00	0.00	0.00
Diseases of the aortic and mitral valves	396x	4,940	98.74	0.97	0.28
Diseases of other endocardial structures	397x	2,657	97.93	1.77	0.30
Other rheumatic diseases	398x	1,901	96.37	3.16	0.47
Anomalies of cardiac septal closure	745x	419	99.05	0.72	0.24
Other congenital anomalies of the heart	746	491	99.80	0.00	0.20
Essential hypertension	401x	321,026	99.35	0.19	0.45
Hypertensive heart disease	402x	24,592	98.29	1.32	0.39
Hypertensive renal disease	403x	18,443	98.05	1.42	0.53
Hypertensive heart and renal disease	404x	4,246	98.00	1.48	0.52
Secondary hypertension	405x	362	94.75	4.42	0.83
Old myocardial infarction	412	52,813	99.19	0.42	0.39
Other forms of chronic ischemic heart disease	414x	412,053	99.29	0.36	0.35
Chronic pulmonary heart disease	416x	6,327	95.08	4.17	0.74
Cardiomyopathy	425x	16,981	98.43	1.03	0.54
Atherosclerosis	440x	19,404	98.33	1.00	0.67
Aortic aneurysm and dissection	441x	13,946	98.67	1.00	0.32
Other peripheral vascular disease	443x	23,807	99.28	0.38	0.34
Chronic bronchitis	491x	12,130	94.09	5.54	0.37
Emphysema	492x	4,509	98.89	0.86	0.24
Asthma	493x	16,244	98.69	0.98	0.33
Bronchiectasis	494x	441	95.92	3.17	0.91

Chronic airway obstruction	496	57,676	99.13	0.42	0.44
Chronic liver disease & cirrhosis	571x	2,619	98.78	0.76	0.46
Chronic pancreatitis	577.1	331	96.98	2.42	0.60
Morbid obesity	278.01	6,066	99.62	0.12	0.26
Chronic glomerulonephritis	582x	718	95.13	0.56	4.32
Chronic renal failure	585	4,109	97.91	1.63	0.46
Multiple sclerosis	340	699	99.57	0.29	0.14
Diffuse disease of connective tissue	710x	2,538	99.01	0.43	0.55
Rheumatoid arthritis	714x	7,121	99.42	0.14	0.44
Disorders of parathyroid gland	252x	483	97.52	1.86	0.62
Disorders of the immune mechanism	279x	107	98.13	0.93	0.93
Hereditary hemolytic anemias	282x	584	98.63	1.20	0.17
Primary thrombocytopenia	287.3	546	91.39	6.96	1.65
<i>Complications</i>					
Functional disturbances following cardiac surgery	429.4	539	21.71	77.55	0.74
Cardiac complications	997.1	22,733	10.58	89.16	0.26
Postoperative shock	998	453	10.38	88.74	0.88
Hemorrhage complicating a procedure	998.11	8,897	12.75	86.73	0.53
Pneumothorax	512x	3,296	16.84	82.65	0.52
Pulmonary insufficiency following trauma or surgery	518.5	6,706	14.35	85.22	0.43
Peripheral vascular complications	997.2	1,329	13.39	85.85	0.75
Respiratory complications	997.3	9,990	10.65	89.13	0.22
Digestive system complications	997.4	3,102	10.77	88.97	0.26
Renal complications	997.5	4,998	10.44	89.02	0.54
CNS complications	997.0	4,107	10.10	89.31	0.58
Postoperative infection	998.5	2,429	17.62	81.93	0.45
Complications of medical care, unclassified	999.0x	1,476	16.67	82.05	1.29

CPAA, conditions present at admission.

Figure 1: Hospital Distribution of Condition Present at Admission (CPAA) Modifier for Chronic Conditions

The y-axis represents the percent of hospitals which coded the CPAA modifier as being present at admission as a function of the percent of chronic conditions coded as present on admission.

patients admitted for a surgical procedure. We then reexamined the hospital distribution of complication codes after excluding patients with more than one admission between 1999 and 2000 to eliminate patients who were coded with complication(s) from prior admissions during their index admission (Figure 3). For most of the hospitals, 90 percent or more of the complications were coded as not being present on admission (Figure 3). In the case of the AAA, CEA, and THR patients, there is good evidence that most complications are coded as not present at admission using the CPAA modifier. For patients undergoing CABG and PTCA, a greater proportion of complication codes are coded as present on admission compared with the AAA, CEA, and THR patient populations. However, it is possible that some of the complications coded as present on admission resulted from outpatient procedures. This is especially likely in both the CABG and PTCA patient populations as many coronary

Figure 2: Hospital Distribution of Condition Present at Admission (CPAA) Modifier for Complication Codes

The y-axis represents the percent of hospitals which coded the CPAA modifier as being present at admission as a function of the percent of complications coded as present on admission.

angioplasties are performed on an outpatient basis. We were not able to test this hypothesis because we did not have access to outpatient data.

Model Development

The dependent variable was in-hospital death (30-day mortality was not available).

Separate hierarchical multivariate logistic regression models were developed for each study population. Hierarchical modeling was used as the outcomes of patients treated at the same hospital may not be independent (Leyland and Goldstein 2001). Each model adjusted for the severity of the principal diagnosis and for comorbid conditions. The severity of the principal diagnosis was coded using the Disease Staging classification system developed

Figure 3: Hospital Distribution of Condition Present at Admission (CPAA) Modifier for Complication Codes after Omitting Patients with Prior Admissions in 1999–2000

The y-axis represents the percent of hospitals which coded the CPAA modifier as being present at admission as a function of the percent of complications coded as present on admission.

by (Gonnella, Louis, and Gozum 1994). Comorbid conditions were coded either using Disease Staging or the Elixhauser algorithm (Elixhauser et al. 1998). The set of Disease Staging categories considered as possible covariates in each model included coronary artery disease, congestive heart failure, cardiomyopathy, valvular disease (aortic stenosis, aortic insufficiency, mitral stenosis, mitral insufficiency), arrhythmias, hypertension, respiratory disease (COPD, emphysema), and endocrine disease (diabetes, adrenal, thyroid). Only conditions from the Elixhauser algorithm that did not overlap with these Disease Staging categories were considered as possible risk factors (i.e., renal failure, liver disease, paralysis, other neurological disease, coagulopathy, fluid and electrolyte disorders, cancer). The set of explanatory variables considered for each model also included demographic variables (age and gender), transfer status, and admission type (elective versus nonelective).

Initial variable selection was performed using standard logistic regression modeling because stepwise procedures for model selection are not available for hierarchical logistic regression models. The method of fractional polynomials by Royston and Altman (1994) was used to determine the optimal transformations for age. These risk factors were then used to construct a random-intercept model using PROC GLIMMIX (Littel et al. 1996) in SAS (SAS Corp., Cary, NC). Model discrimination was evaluated using the C-statistic, which is equivalent to the area under the ROC curve (Hanley and McNeil 1982). Model calibration was not evaluated using the Hosmer–Lemeshow statistic since this measure of model fit should only be used if the observations are independent. Instead, calibration curves were constructed to investigate model calibration.

For each study population, one model was constructed using only secondary diagnoses present at admission based on the CPAA modifier: “date stamp” model. All other secondary diagnoses were excluded from this model. The second model was constructed using all secondary diagnoses, ignoring the information present in the CPAA modifier: the “no date stamp model.”

Identification of Hospital Quality Outliers

Hospital quality outliers were identified using the “date stamp” and the “no date stamp” models. Low- and high-performance hospitals were identified using the ratio of the observed mortality rate to the expected mortality rate (OE ratio) (Shwartz, Ash, and Iezzoni 1997). The expected mortality rate was obtained using the estimated models. Hospitals whose OE ratio was < 1 and whose 95 percent confidence interval (CI) did not include 1 were designated as high-performance hospitals. Similarly, hospitals whose OE ratio was > 1 and whose 95 percent CI did not include 1 were designated as low-performance hospitals. As the hospital random intercept is the hospital-specific contribution to outcome, only the fixed-effects coefficients in the hierarchical model were used to calculate the expected mortality rate (Glance et al. 2003).

Bootstrapping was used to calculate the 95 percent CI around the OE ratio for each hospital using an approach similar to one described by Hosmer and Lemeshow (1995). To incorporate the uncertainty of the model coefficients, the parameter estimates used to predict the mortality rate in each bootstrap sample were randomly drawn from the parameter distribution specified by the estimated values for the fixed-effects parameters and the estimated variance–covariance matrix. One thousand replications were used

to construct the bootstrap confidence intervals using the percentile method (Efron and Tibshirani 1993).

Analysis

For the purposes of the analysis, hospitals identified as quality outliers using the “date stamp” model were considered “true” outliers. We assessed the accuracy of the no-date stamp model for identifying hospital quality outliers using the findings based on the “date stamp” model as the “gold” standard and the findings of the “no date stamp” model as the “test” being evaluated.

For each cohort (e.g., CABG population), we calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the “no date stamp” model for identifying low-performing hospitals and high-performing hospitals.

We also performed an analysis to determine whether hospitals whose outlier status changed (e.g., low quality to average quality) after using the “date stamp” model were also hospitals which were either more or less likely to code complications as preexisting conditions—compared with hospitals whose outlier status was unchanged. If systematic miscoding of the CPAA modifier were responsible for changes in hospital quality noted with date stamping, then hospitals which coded more complications as present at admission would be more likely to be reclassified as higher quality hospitals using the date stamp models compared with the “no date stamp” models: coding complications as present at admission would increase the apparent severity of disease and increase the hospital’s predicted mortality rate. Conversely, hospitals which coded fewer complications as present at admission would be more likely to be reclassified as lower quality outliers.

To test this hypothesis, we first divided the hospitals into three groups: (1) hospitals which were reclassified as higher quality using the “date stamp” models; (2) hospitals whose quality remain unchanged; and (3) hospitals which were reclassified as lower quality using the “date stamp” models. We then constructed separate logistic regression models for five of the study populations (AAA, CABG, PTCA, CEA, and THR). We excluded AMI and stroke patients from this analysis as we would expect more heterogeneity in the application of the CPAA modifier to “complications” in these patient populations because of hospital readmissions. The dependent variable was the CPAA modifier (0/1) for ICD-9-CM complication codes and the explanatory variable was the hospital group. Robust variance estimators were used to account for the lack of independence of CPAA modifier coding within hos-

Table 3: Demographics of Study Populations

	<i>CABG</i>	<i>PTCA</i>	<i>CEA</i>	<i>AAA</i>	<i>THR</i>	<i>AMI</i>	<i>CVA</i>
Age, mean (SD)	66 (11)	65 (12)	72 (9)	73 (8)	72 (14)	69 (14)	73 (13)
Male (%)	72	67	57	79	35	61	44
Number of diagnoses, median (IQR)	6 (4,8)	5 (3,7)	5 (3,6)	6 (4,8)	4 (2, 6)	6 (4,8)	6 (5,9)
Mortality rate (%)	3.0	1.6	0.88	12.5	1.13	8.9	6.4
Hospital volume,* median (IQR)	509 (298,826)	734 (68,1514)	65 (20,167)	17 (6,44)	131 (57,287)	319 (117,665)	254 (106,482)

*Hospital volume calculated for period of 1998–2000.
SD, standard deviation; IQR, interquartile range; CABG, coronary artery bypass graft surgery; PTCA, coronary angioplasty; CEA, carotidendarterectomy; AAA, abdominal aortic aneurysm surgery; THR, total hip replacement; AMI, acute myocardial infarction; CVA, stroke.

pitals. We limited this analysis to complication codes as: (1) there was virtually no variability in how hospitals applied the CPAA modifier to chronic conditions and (2) including ICD-9-CM codes which could either represent pre-existing conditions or complications in this analysis would necessarily show that the application of the CPAA modifier influenced hospital quality assignment—otherwise, the CPAA modifier would convey no real information.

All statistical analyses were performed using *STATA SE* versions 8.2 and 9.0 (STATA Corp., College Station, TX) and *SAS* release 8.02 (SAS Corp).

RESULTS

This study was based on seven study populations and included 648,866 patients (Table 1). This demographic information is summarized in Table 3. The

Table 4: Performance of Models Predicting In-Hospital Mortality

	<i>C-Statistic</i>	
	<i>No Date Stamp Model</i>	<i>Date Stamp Model</i>
Coronary artery bypass grafting	0.92	0.81
Coronary angioplasty	0.95	0.88
Carotid endarterectomy	0.93	0.84
Abdominal aortic aneurysm repair	0.92	0.88
Hip replacement	0.94	0.88
Acute myocardial infarction	0.89	0.84
Stroke	0.79	0.79

study populations were predominantly male and elderly (age > 65). The in-hospital mortality rate varied between 0.88 percent for patients undergoing CEA to 12.5 percent for patients undergoing AAA repair. The median hospital volumes over the 3-year study period varied greatly as a function of the medical condition: from 17 for AAA repair to 734 for PTCA. The median number of diagnostic codes varied between four and six for each of the study populations.

All models exhibited excellent discrimination (C-statistic ≥ 0.79) (Table 4). Except for the stroke models, all of the “no date stamp” models exhibited better discrimination than the “date stamp” models. The increase in discriminatory power resulting from miscoding complications as preexisting conditions has been previously reported (Hannan et al. 1997). As patients with complications are at greater risk for dying, mortality prediction models that include complications would necessarily predict mortality more accurately than models that do not include complications. Inspection of model calibration plots suggests that model calibration was very good for all of the models. The calibration plots for the CABG models are shown in the Appendix 1. (Other calibration plots are similar, and are available upon request.)

Figure 4 displays the OE ratio for the CABG hospitals. To improve clarity, only the hospitals for which the “date stamp” and the “no date stamp” models assigned different outlier status are displayed. For each hospital, the point estimate and the 95 percent confidence interval band for the OE ratio are shown. The “date stamp” model identified 34 quality outlier CABG hospitals and the “no date stamp” model identified 29 quality outlier CABG hospitals. The two models disagreed on the outlier status of 25 hospitals out of a total of 123 hospitals performing CABG surgery. Results for the other six patient populations are available on request.

The extent to which hospitals were misclassified as quality outliers using the “no date stamp” models is shown in Tables 5 and 6. The proportion of low-quality hospitals (identified using the “date stamp” models) which were “missed” by the “no date stamp” models is equal to one minus the sensitivity of the “no date stamp” models (defined as the false-negative error rate). Forty percent of the CABG hospitals identified as low-performance hospitals by the “date stamp” model were not classified as low-performance hospitals by the “no date stamp” model. Similarly, 33 percent of the low-performance PTCA hospitals, 0 percent of the low-performance CEA hospitals, 6 percent of the low-performance AAA repair hospitals, 40 percent of the low-performance hip replacement hospitals, 33 percent of the low-performance AMI hospitals, and 8 percent of the low-performance stroke

Figure 4: Standardized Mortality Ratio (OMR/EMR) for Coronary Artery Bypass Graft—Impact of Date Stamping of ICD-9-CM Codes on Identity of Low-Performance and High-Performance Hospitals

Hospitals with an OMR/EMR ratio less than 1 and with a 95% confidence interval that does not cross 1 are high-quality outliers. Hospitals with an OMR/EMR ratio greater than 1 and with a 95% confidence interval that does not cross 1 are low-quality outliers. Each hospital SMR is represented twice: first based on the “date stamp” model (light gray), then based on the “no date stamp” model (dark gray).
*indicates that OMR/EMR was obtained using the date-stamp model abbreviations: OMR, observed mortality rate; EMR, expected mortality rate

Table 5: Identification of Low-Performance Hospitals by “No Date Stamp” Models

<i>Condition</i>	<i>n</i>	<i>Sensitivity</i>	<i>FNER</i>	<i>PPV</i>	<i>FPER</i>
Coronary artery bypass	123	0.60	0.40	0.50	0.50
Coronary angioplasty	162	0.67	0.33	0.67	0.33
Carotid endarterectomy	311	1.00	0	0.50	0.50
Abdominal aortic aneurysm repair	301	0.94	0.06	0.85	0.15
Hip replacement	360	0.60	0.40	0.75	0.25
Acute myocardial infarction	378	0.67	0.33	0.64	0.36
Stroke	389	0.92	0.08	0.88	0.12

n, number of hospitals; FNER, false-negative error rate (proportion of low-performance hospitals “missed” by the “no date stamp” models); PPV, positive predictive value; FPER, false-positive error rate (proportion of hospitals misclassified as low-quality hospitals by “no date stamp” model).

hospitals were not classified as low-performance hospitals by the “no date stamp” model.

The proportion of hospitals identified as low-quality hospitals by the “no date stamp” models but not by the “date stamp” models is one minus the PPV of the “no date stamp” models (defined as the false-positive error rate). Half of the hospitals identified as low-performance CABG hospitals by the “no date stamp” model were not identified as low-performance hospitals by the “date stamp” model. Similarly, 33 percent of the PTCA hospitals, 50 percent of the CEA hospitals, 15 percent of the AAA repair hospitals, 25 percent of the hip replacement hospitals, 36 percent of the AMI hospitals, and 12 percent of the

Table 6: Identification of High Performance Hospitals by “No Date Stamp” Models

<i>Condition</i>	<i>n</i>	<i>Sensitivity</i>	<i>FNER</i>	<i>PPV</i>	<i>FPER</i>
Coronary artery bypass	123	0.54	0.46	0.76	0.24
Coronary angioplasty	162	0.82	0.18	0.84	0.16
Carotid endarterectomy	311	0.99	0.01	0.99	0.01
Abdominal aortic aneurysm repair	301	0.97	0.03	0.87	0.13
Hip replacement	360	0.96	0.04	0.97	0.03
Acute myocardial infarction	378	0.85	0.15	0.77	0.23
Stroke	389	0.95	0.05	1.00	0

n, number of hospitals; FNER, false-negative error rate (proportion of high-performance hospitals “missed” by the “no date stamp” models); PPV, positive predictive value; FPER, false-positive error rate (proportion of hospitals misclassified as high-quality hospitals by “no date stamp” model).

stroke hospitals identified as low-performance hospitals by the “no date stamp” models were not identified as low-performance hospitals by the “date stamp” models.

The misclassification rates for high-performance hospitals were calculated in a similar fashion. In general, the level of agreement between the “date stamp” model and the “no date stamp” models was better with regards to identifying high-performance hospitals than it was for identifying low-performance hospitals. These results are shown in Table 6.

Finally, there was no statistically significant association between changes in hospital quality assessment (“date stamp” versus “no date stamp” models) and whether or not hospitals were more or less likely to code complications as present at admission using the CPAA modifier (Table 7). However, in the angioplasty group, hospitals which were reclassified as being higher quality using the “date stamp” model tended to code a higher proportion of complications as present at admission ($OR = 1.51, p = .056$).

DISCUSSION

Quality report cards may have the potential to significantly improve health care quality. Report cards can be used to identify high performance hospitals and these hospitals, in turn, can be studied to identify best practices (Epstein 1995). These best practices can then be disseminated to other hospitals. Report cards can also be used to steer patients toward high-performance hospitals (selective referral) and away from low-performance hospitals (selective avoidance). However, performance profiling is not without risk. In a seminal paper entitled the “Risks of Risk Adjustment,” Iezzoni (1997a) highlighted one of the fundamental problems of performance profiling: whether or not a hospital is identified as a low-performance or high-performance hospital depends substantially on which scoring system is used to construct the quality ranking. Therein lies one of the major “risks” of report cards: using them as the basis for evidence-based patient referrals may not necessarily lead to better population outcomes if the hospital rankings are not accurate. A priori, report cards based on prediction models that do not distinguish between preexisting conditions and complications would be expected to lead to less accurate quality assessment.

Because administrative data are widely available and relatively inexpensive, they are being increasingly used by third-party payers and the States to construct hospital and physician report cards. However, administrative data

Table 7: Analysis of the Association between Change In-Hospital Quality Assessment and the Proportion of “Complications” Date Stamped as Being Present on Admission

	AAA			CABG			CEA			PTCA			THR		
	OR	p-Value		OR	p-Value		OR	p-Value		OR	p-Value		OR	p-Value	
Change to higher quality	1.64	.39		1.20	.46		1.88	.15		1.51	.056		0.77	.48	
No change in outlier status	Reference			Reference			Reference			Reference			Reference		
Change to lower quality	1.66	.18		1.64	.44		0.86	.18		0.72	.37		0.52	.35	

An odds ratio > 1 indicates that “complications” were more likely to be coded as being present on admission compared to the reference population. An odds ratio < 1 indicates that “complications” were less likely to be coded as being present on admission compared to the reference population. Assuming that misclassifying “complications” as being present on admission accounted for hospitals being reclassified as higher quality outliers using risk adjustment models incorporating date stamp information, then those hospitals would be expected to have an odds ratio > 1.

CABG, coronary artery bypass graft surgery; PTCA, coronary angioplasty; CEA, carotidendarterectomy; AAA, abdominal aortic aneurysm surgery; THR, total hip replacement.

have a major limitation. Most ICD-9-CM codes do not indicate whether a secondary diagnosis represents a preexisting condition or a complication. We hypothesized that the absence of date stamping in administrative data, by preventing the accurate specification of disease severity and comorbidities, would yield biased measures of hospital quality.

This study shows that risk-adjustment models based on routine administrative data frequently misidentified low-quality and high-quality hospitals. For example, one-half of the hospitals identified as low-quality CABG hospitals should not have been classified as low-quality outliers. Forty percent of the hospitals providing low-quality care to patients undergoing THR and one-third of the hospitals providing low-quality care to patients admitted with acute myocardial infarction were “missed” using risk-adjustment models based on routine administrative data that did not include date stamp information. We found significant misclassification rates for four out of the seven conditions included in this study: CABG surgery, PTCA, hip replacement, and AMI.

To our knowledge, only two other studies have assessed the impact of date-stamping ICD-9-CM codes in administrative data on the accuracy of quality report cards. Romano and Chan (2000) studied the impact of using only conditions present at admission to measure hospital quality in a stratified sample of 974 patients with AMIs. Chart abstraction was used to identify preexisting conditions. They found that using only conditions present at admission had a “major impact on the classification of hospitals [quality] based on risk-adjusted mortality.” A study by Ghali, Quan and Bran (2001), based on 50,397 patients undergoing CABG surgery at 23 Canadian hospitals over a 4-year period, showed that the use of a diagnosis-type indicator, similar to the CPAA in California, also had a significant impact on hospital ranking. The findings of our study corroborate the earlier work by Romano and Ghali and further extends their work to include other patient populations.

Clinical data are superior to administrative data for measuring quality. The limitations of administrative data have been well described: errors in the abstraction process, undercoding of comorbidities (Jollis et al. 1993), variation in the quality of charting by physicians (Pronovost and Angus 1999), lack of precise definitions for ICD-9-CM codes (Iezzoni 1997b), underreporting of “complication” codes (Jencks 1992), and “overcoding” of patient diagnoses to maximize reimbursements (Iezzoni 1997c). But despite the limitations of administrative data, large-scale efforts to measure quality will continue to rely primarily on administrative data until the necessary information technology infrastructure is created to collect clinical data in a routine and cost-effective

fashion. Furthermore, clinical datasets also present problems. The possibility of “gaming” the data always exists when clinicians are responsible for collecting data on their own patients. Because administrative datasets seem destined in the near term to be the foundation of quality assessment, it is imperative that these datasets be as robust as possible. The findings of this study strongly support the addition of CPAA modifiers to indicate whether an ICD-9-CM code describes a secondary diagnosis present at the time of hospital admission.

This study has several strengths. This is the first major study to evaluate the impact of date stamping administrative data on hospital quality reporting across a large spectrum of surgical and medical conditions. Second, the California OSHPD routinely monitors data quality of the California SID. Discharge data reports that do not meet error tolerance levels established by the state are sent back to the reporting institution for correction (California Patient Discharge Data Reporting Manual 2000). Third, the scope and size of this population-based study, and the fact that it did not exclude patients younger than 65, makes it reasonable to assume that the study conclusions can be generalized to include other medical conditions, as well as populations outside of California.

This study has several important limitations. First, we have assumed that the CPAA modifier accurately identifies conditions present at admission. This data field has not been validated using either chart abstraction or clinical data. However, an exploratory analysis suggests that this data field has construct validity. For ICD-9-CM codes that are very likely to code chronic conditions, the CPAA modifier indicated that these secondary diagnoses were present at admission 99 percent of the time. For ICD-9-CM codes that designate secondary diagnoses likely to represent complications, the CPAA modifier indicated that these secondary diagnoses were present at admission in 12 percent of the cases. Although the CPAA modifier is being uniformly applied by hospitals to code chronic conditions, there is some heterogeneity in the manner in which hospitals code the CPAA modifier for ICD-9-CM complications codes. It appears that some of this heterogeneity is secondary to complications from previous admissions being correctly coded as present on admission.

However, we did not directly evaluate the validity of the CPPA modifier for ICD-9-CM codes that could represent either preexisting conditions or complications (i.e., AMI). Conditions coded by these ICD-9-CM codes are the very codes for which accurate coding of the CPAA modifier is necessary to assure that hospitals do not receive “credit” for their complications. Such an evaluation of the CPAA modifier would require chart abstraction. Our

analyses of the CPAA modifier were based on the assumption that the degree of accuracy with which hospitals coded the CPAA modifier for either chronic conditions (i.e., COPD) or complications (i.e., postoperative shock) could be extrapolated to conditions which could be either chronic conditions or complications. Therefore, our analysis of the validity of the CPAA modifier provides indirect evidence of the validity of the CPAA modifier and awaits further validation using systematic chart review.

Importantly, we did not find a significant association between hospital coding of complications and changes in hospital quality with date stamping. It is therefore unlikely that miscoding of the CPAA modifier at the hospital level would have accounted for most of the changes in hospital quality assessment that we observed with the application of date stamp information. However, this analysis was based on a limited number of ICD-9-CM codes that were chosen because they would be expected to code for complications as opposed to preexisting conditions. The lack of an association between the coding of the CPAA modifier for complications and changes in quality assessment provides indirect support for the hypothesis that miscoding of the CPAA modifier, in general, did not lead to the changes in quality outlier status that were observed in this study. We did not evaluate the possible association between miscoding of the CPAA modifier for ICD-9-CM codes which could represent either preexisting conditions or complications because, short of conducting a chart audit, we had no way to verify whether the CPAA modifier was being correctly coded in those cases. Therefore, as we did not perform chart reabstraction, and only explored the association of the CPAA modifier for ICD-9-CM complication codes with changes in hospital quality, we cannot rule out the possibility that systematic biases in the coding of the CPAA modifier for other conditions (which could either represent preexisting conditions or complications) could have led to the finding that the absence of date stamp information occasionally leads to the misclassification of hospital quality outliers.

Second, we have defined the “date stamp” models as the “gold standard” for identifying low-performance and high-performance hospitals. While there is no perfect way of identifying “true” quality outliers, and while the “date stamp” has construct validity and offers an opportunity to improve the distinction between preexisting conditions and complications, it requires further validation. Nevertheless, the large substantive differences in the identification of outliers between the “date stamp” and the “no date stamp” models show the importance of identifying complications before constructing risk-adjusted quality measures, and calls into question the use of administrative data for these purposes in the absence of date stamp information.

Third, the definition of quality outliers as hospitals whose observed-to-expected mortality (OE) ratio is statistically different from “1” will result in some hospitals being classified differently as a result of relatively small changes in the confidence interval for the OE ratio. Small quantitative changes in a numerical scale (OE ratio 95 percent confidence interval) can result in large qualitative changes in a categorical scale (low- and high-performance hospital). For example, a hospital will be reclassified as a low performance hospital if the 95 percent confidence interval around its OE ratio changes from 0.98–1.10 to 1.01–1.12. Despite the limitations of using confidence intervals to identify quality outliers, this approach is preferable to simply using point estimates of the OE ratio alone without conveying the extent of statistical uncertainty around the point estimate of the OE ratio.

Fourth, in-hospital mortality was used instead of 30-day mortality as the outcome of interest because information of 30-day mortality was not available in our dataset. Although differences in discharge policies could create a bias when comparing quality across different hospitals, the use of in-hospital mortality would not be expected to bias the study findings as each hospital was compared with itself. Finally, mortality is one of several important dimensions of health care quality. We chose to study mortality as the outcome of interest because most current efforts to evaluate health care quality using outcomes measurement are based on mortality, and do not include other important outcomes such as functional status.

CONCLUSION

This study supports the hypothesis that the use of routine administrative data without date stamp information to construct hospital quality report cards may result in the mis-identification of quality outliers. Until report cards based on clinical data become more widely available, we recommend that CPAA modifiers be added to all administrative data sets. However, the CPAA modifier will need to be further validated before date stamped administrative data can be used as the basis for health quality report cards.

ACKNOWLEDGMENT

This project was supported by a grant from the Agency for Healthcare and Quality Research (RO1 HS 13617).

NOTES

1. Of the 12 conditions studied, the strength of agreement between the CPAA field and reabstracted data was poor ($n = 1$), slight ($n = 3$), fair ($n = 4$), moderate ($n = 3$), and substantial ($n = 1$).
2. "While the OSHPD CPAA indicator has not yet undergone a broad validation through medical chart reabstraction, several internal analyses give us considerable confidence in its general validity. These include trending analyses for the coding of chronic conditions and complications where improvements over time have been noted. Linked administrative-clinical data analyses in the area of cardiac surgery and comparisons of disease prevalence in selected disease/surgical cohorts with those in the clinical literature have provided additional evidence of coding accuracy. The 1996 community acquired pneumonia validation study found that only three out of 13 nonchronic conditions (all candidate risk factors) exhibited κ coefficients of 0.47 or better when OSHPD's CPAA data was compared with the 'gold standard' medical chart. However, that study was done the year of the CPAA's introduction and trending analysis suggests that CPAA coding accuracy has improved greatly since 1996" (Joseph P. Parker, Ph.D., Director, Healthcare Outcomes Center, Office of [California] Statewide Health Planning and Development, February, 2005, personal communication).

REFERENCES

- Blumenthal, D. 1996. "Quality of Health Care. Part 4: The Origins of the Quality-of-Care Debate." *New England Journal of Medicine* 335: 1146–49.
- California Patient Discharge Data Reporting Manual. 2000. *Office of Statewide Health Planning and Development*. Sacramento, CA: Government of CA.
- Chassin, M. R., and R. W. Galvin. 1998. "The Urgent Need to Improve Health Care Quality. Institute of Medicine National Roundtable on Health Care Quality." *Journal of American Medical Association* 280: 1000–5.
- Committee on Quality of Health Care in America 2000. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press.
- . 2001. *Crossing the Quality Chasm*. Washington, DC: National Academy Press.
- Efron, F., and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey. 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36: 8–27.
- Epstein, A. 1995. "Performance Reports on Quality—Prototypes, Problems, and Prospects." *New England Journal of Medicine* 333: 57–61.
- Fernandopulle, R., T. Ferris, A. Epstein, B. McNeil, J. Newhouse, G. Pisano, and D. Blumenthal. 2003. "A Research Agenda for Bridging the 'Quality Chasm.'" *Health Affairs (Millwood)* 22: 178–90.

- Ghali, W. A., H. Quan, and R. Brant. 2001. "Risk Adjustment Using Administrative Data: Impact of a Diagnosis-Type Indicator." *Journal of General Internal Medicine* 16: 519-24.
- Glance, L. G., A. W. Dick, T. M. Osler, and D. Mukamel. 2003. "Using Hierarchical Modeling to Measure ICU Quality." *Intensive Care Medicine* 29: 2223-29.
- Glance, L. G., A. W. Dick, T. M. Osler, and D. B. Mukamel. 2006. "Does Date Stamping ICD-9-CM Codes Increase the Value of Clinical Information in Administrative Data?" *Health Services Research* 41(1): 231-51.
- Gonnella, J. S., D. Z. Louis, and M. E. Gozum. 1994. *Disease Staging: Classification and Severity Stratification*. Ann Arbor, MI: The Medstat Group.
- Hanley, J. A., and B. J. McNeil. 1982. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." *Radiology* 143: 29-36.
- Hannan, E. L., M. J. Racz, J. G. Jollis, and E. D. Peterson. 1997. "Using Medicare Claims Data to Assess Provider Quality for CABG Surgery: Does It Work Well Enough?" *Health Services Research* 31: 659-78.
- Hersh, W. 2004. "Health Care Information Technology: Progress and Barriers." *Journal of the American Medical Association* 292: 2273-74.
- Hosmer, D. W., and S. Lemeshow. 1995. "Confidence Interval Estimates of an Index of Quality Performance Based on Logistic Regression Models." *Statistics in Medicine* 14: 2161-72.
- Hurtado, M. P., E. K. Swift, and J. M. Corrigan. 2001. *Envisioning the National Health Care Quality Report*. Washington, DC: National Academy Press.
- Iezzoni, L. I. 1997a. "The Risks of Risk Adjustment." *Journal of the American Medical Association* 278: 1600-7.
- . 1997b. "Assessing Quality Using Administrative Data." *Annals of Internal Medicine* 127: 666-74.
- . 1997c. "Assessing Quality Using Administrative Data." *Annals of Internal Medicine* 127: 666-74.
- Jencks, S. F. 1992. "Accuracy in Recorded Diagnoses." *Journal of the American Medical Association* 267: 2238-39.
- Jollis, J. G., M. Ancukiewicz, E. R. DeLong, D. B. Pryor, L. H. Muhlbaier, and D. B. Mark. 1993. "Discordance of Databases Designed for Claims Payment versus Clinical Information Systems. Implications for Outcomes Research." *Annals of Internal Medicine* 119: 844-50.
- Jollis, J. G., and P. S. Romano. 1998. "Pennsylvania's Focus on Heart Attack—Grading the Scorecard." *New England Journal of Medicine* 338: 983-87.
- Leatherman, S. T., J. H. Hibbard, and E. A. McGlynn. 2003. "A Research Agenda to Advance Quality Measurement and Improvement." *Medical Care* 41: 180-6.
- Leyland, A. H., and H. Goldstein. 2001. *Multilevel Modelling of Health Statistics*. New York: John Wiley & Sons Ltd.
- Littel, R., G. A. Millikan, W. W. Stroup, and R. D. Wolfinger. 1996. *SAS System for Mixed Models*. Cary, NC: SAS Institute.
- McGlynn, E. A. 2003. "An Evidence-Based National Quality Measurement and Reporting System." *Medical Care* 41: 18-15.

- McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. 2003. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348: 2635–45.
- McGlynn, E. A., and R. H. Brook. 2001. "Keeping Quality on the Policy Agenda." *Health Affairs (Millwood)* 20: 82–90.
- McGlynn, E. A., C. K. Cassel, S. T. Leatherman, A. DeCristofaro, and H. L. Smits. 2003. "Establishing National Goals for Quality Improvement." *Medical Care* 41: I16–29.
- Pronovost, P., and D. C. Angus. 1999. "Using Large-Scale Databases to Measure Outcomes in Critical Care." *Critical Care Clinics* 15: 615–viii.
- Report for the California Hospital Outcomes Project Community-Acquired Pneumonia, 1996: Model Development and Validation, OSHPD, 2000.
- Romano, P. S., and B. K. Chan. 2000. "Risk-Adjusting Acute Myocardial Infarction Mortality: Are APR-DRGs the Right Tool?" *Health Services Research* 34: 1469–89.
- Royston, P., and D. G. Altman. 1994. "Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modeling." *Applied Statistics* 43: 429–67.
- Shaller, D., S. Sofaer, S. D. Findlay, J. H. Hibbard, D. Lansky, and S. Delbanco. 2003. "Consumers and Quality-Driven Health Care: A Call to Action." *Health Affairs (Millwood)* 22: 95–101.
- Shwartz, M., A. S. Ash, and L. I. Iezzoni. 1997. "Comparing Outcomes Across Providers." In *Risk Adjustment for Measuring Healthcare Outcomes*. 2nd edition, edited by L. I. Iezzoni, pp. 471–516. Chicago: AHSR.

SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available online:

APPENDIX 1. Calibration Plots for the CABG Models.